Heterogeneity and Side Deals in Climate Negotiations: Evidence from a Bargaining Experiment

Greer Gosnell[1, 2], Alessandro Tavoni[2, a]

[1] Geography and Environment Department, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE.
[2] Grantham Research Institute, London School of Economics and Political Science, Houghton Street, London, WC2A 2AE.
[a]To whom correspondence should be addressed. Email: a.tavoni@lse.ac.uk.

Abstract: Given the continual rise in global CO_2 emissions, the current state of affairs in international climate negotiations provides little reason for optimism. The UNFCCC approach to seeking universal participation has thus been called into question, both by policy makers and by academics, who have established pessimistic theoretical predictions concerning the ability of international environmental agreements to improve upon nation states' policy decisions in the absence of such an agreement. Focusing on variations of the public goods game, game theorists have predicted that self-enforcing agreements are likely to comprise only a handful of countries committing to unambitious emissions abatement targets. We focus instead on the dynamics of the negotiation process by studying experimental behavior in a Nash bargaining game involving a six-player group of negotiators representing heterogeneous countries. Throughout repeated rounds of negotiation, subjects bargain over the allocation of a fixed amount of profit-generating emissions. Negotiating groups collectively determine whether the global emissions reduction target is reached, and there are significant losses associated with prolonged failure to reach an agreement. The treatments focus on wealth (and responsibility) asymmetry, as well as on the potential of preliminary side agreements among homogeneous subsets of players to ease coordination of demands in keeping with the target. While these side agreements do not ensure success, they alter the dynamics of the bargaining process considerably and differently for Poor and Rich negotiators. We attribute these shifts in dynamics to the pronounced salience of inequities under the various contexts posed in our treatments.

Significance statement: Consistent with UNFCCC progress to date, academic research has historically cast doubt upon existing international approaches to achieving universal commitment on greenhouse gas emissions abatement, establishing pessimistic predictions concerning the ability of international agreements to improve upon business-as-usual. We introduce a new experimental framework, focusing on the dynamics of the negotiation process by observing experimental behavior in a framed Nash bargaining game involving a six-player group of negotiators representing heterogeneous countries. Throughout repeated negotiation rounds, subjects bargain over the allocation of a fixed amount of profit-generating emissions with significant losses attached to prolonged failure to reach agreement. We find that side agreements do not ensure success, but that they alter the dynamics of the negotiation process considerably.

INTRODUCTION

Recent developments in climate policy have reaffirmed the perceived importance of minilateral agreements among a small number of countries prior to engaging in large *fora* such as the upcoming Paris Conference of the Parties (COP). A growing literature, notably in international relations and political science, points to the merits and drawbacks of entering into negotiations among small-n clubs (1-3). At the two ends of the spectrum, one finds bilateral negotiations and almost universal groupings like the UNFCCC COPs. Most agree that bottom-up and top-down approaches are not mutually exclusive (4, 5). Indeed, it appears that some countries have resorted to bilateral deals as a stimulus for action by less motivated countries, a common reading of the U.S.-China agreement to reduce emissions that took place ahead of the Paris meeting in November 2014 and is so far confirmed by the Intended Nationally Determined Contribution (INDC) recently pledged by both countries (6).

Will more reluctant countries commit to emissions cuts once assured of others' intentions to invest in climate change mitigation? This question is of course an empirical one, and the outcome of the Paris COP in December 2015 will provide an indication of whether such assurance matters. In the meantime, one may approach the issue with other tools, such as theoretical modeling and laboratory experimentation. Inspired by a bargaining model that aims to capture some of the stylized tradeoffs inherent in climate change negotiations (7), we introduce a novel economic experiment that focuses on the role of side deals reached by a subset of negotiators in driving behavior in the subsequent global negotiations.

Smead and coauthors (7) use an agent-based model with learning dynamics to examine past failures and future prospects for an international climate agreement. In the model, agents play an N-player Nash bargaining game (8, 9), where each player's strategy set is the interval [0,1] representing the range of possible reductions: 1 representing business-as-usual (BAU)—i.e. 100% emissions—and 0 representing a complete reduction to zero emissions. In addition to imposing learning dynamics, they modify the Nash bargaining game by introducing a global emissions target T in the interval (0,1). Players maintain the full amount demanded from the shared resource or "emissions pie"—where a higher share translates to a higher payoff—only if the sum of all individual demands does not exceed T, which they fix at 50% of BAU emissions. To mimic the cost of failing to reach an agreement, the players' incomes are equal to a fraction δ of their respective demands if the aggregate demand is above the threshold.

The authors vary a number of parameters in the model and find that, unsurprisingly, negotiations are more likely to be successful—i.e. players are more likely to converge on a set of demands consistent with the threshold—as δ (the "disvalue of failure") increases. They also find that prior minilateral agreements can facilitate agreement, especially those made among a large number of small players relative to those made among a small number of large players accounting for the same proportion of emissions.

We explore the issue of negotiating on costly emissions reductions in the laboratory. The experimental literature on the avoidance of dangerous climate change has thus far focused on the provision of threshold public goods (10-15). The underlying idea is that, in order to stay within a safe operating space and avoid probabilistic losses arising from crossing a tipping point for dangerous climate change, players must invest sufficient resources into a public account (16-21).

One can view this public good as a minimum collective expenditure in climate change mitigation that ensures staying below an agreed temperature change, such as the often-mentioned 2°C target.

We instead frame the costly mitigation problem as a Nash bargaining game. Negotiators must divide the burden of reducing the size of the emissions pie by agreeing on sufficiently ambitious reductions relative to BAU, which in the game is represented by players' initial endowments. The underlying assumption is that emissions map one-to-one with wealth. While this assumption is undoubtedly a strong simplification of complex dynamics, it allows us to isolate important features of climate change negotiations, such as the tension between a country's incentive to keep the largest possible fraction of its emissions and the need to make concessions if the collective target is to be met. Furthermore, while historical responsibilities are not specifically modeled here, our main treatments feature wealth heterogeneity so as to capture the different implications of a given reduction commitment (e.g., -20% emissions) by rich, high-emitting countries relative to poor, low-emitting ones.

In addition to the experimental methodology employed, we depart from (7) in several ways. First, in our design, the loss incurred by a group that fails to reach agreement is independent of individual demands (see Materials and Methods). This feature is consistent with the standard bargaining game formulation that prescribes that out-of-equilibrium payoffs are constant. In (7), the out-of-equilibrium payoff is proportional to the player's demand, so that even when agreement is not reached, a player earns more when demanding more.

Perhaps more importantly, to capture the realistic feature that delay in reaching agreement over ambitious emissions reductions will result in the need to agree on even more ambitious targets in the future, we designed the game to comprise multiple rounds with increasingly stringent targets (see Materials and Methods and the Supporting Information (SI)). Hence, while selfish motives still push in the direction of high demands in the hope that others will lead the effort, there is a critical urgency for the negotiating group to meet its target. Reaching agreement early is much less costly than reaching agreement toward the end of the game (which is still preferable to not reaching an agreement at all). Given these features of the experiment, we look at various measures of success to capture the element of velocity in converging to T (see Results).

MATERIALS AND METHODS

Game. Our main treatments focus on asymmetric countries negotiating over a maximum of eight rounds on increasingly ambitious emissions reduction targets. In this process, four Poor Countries and two Rich Countries make successive demands relative to BAU.

Each treatment consists of up to eight rounds of a Nash bargaining game framed as a climate change negotiation, where the negotiation terminates if the group meets the prescribed Global Target T in a given round. The Global Target becomes more difficult to attain as the game progresses, beginning at 60% of global wealth and reducing by 10% every two rounds. If the group does not meet the target by the end of Round 8, group members receive 10% of their initial endowment (regardless of their demands in the final round) as an unavoidable consequence of "dangerous" climate change.

In every round, group members—each acting as a delegate representing one country in the negotiation—engage in what we term the Global Negotiation stage. In this stage, each delegate

demands to keep a proportion of her country's endowed wealth, which is perfectly correlated with her country's emissions in the game. If the group's total demand does not exceed the corresponding Global Target in a given round, the target is met and each subject in the group receives the proportion she demanded in that round. If the target is not met, there is no payout and negotiations continue to the next round.

We implement five variants of the bargaining experiment: Symmetric (SYM), Asymmetric (ASYM), Poor Side Deals (PSD), Rich Side Deals (RSD), and All Side Deals (ASD). All group's aggregate endowments are £100 (approximately US$156). In treatment SYM, all Countries begin with a symmetric endowment of £16.67. All other treatments are characterized by asymmetry in the distribution of endowments (and corresponding impact on global CO_2 emissions). In these treatments, four Poor Country delegates receive an endowment of £10 and two Rich Country delegates receive an endowment of £30.

While different in terms of endowment, both SYM and ASYM feature eight single-stage rounds, as depicted in Figure 1. In each of these rounds, delegates independently and simultaneously decide on individual (i.e. country-level) demands. The software computes the demand of the group and displays both group and individual demands in a subsequent screen in absolute and percentage terms. In treatments containing Side Deals (PSD, RSD, and ASD), a subset(s) of delegates may collectively place binding constraints on the their demands in the upcoming Global Negotiation stages on a given target. Accordingly, these Side Deals take place prior to Rounds 1, 3, 5, and 7.

In each Side Deal, delegates representing Poor Countries (in PSD), those representing Rich Countries (in RSD), or both subgroups simultaneously (in ASD) implement a binding upper bound on the amount of individual wealth that each of them may demand in the upcoming two Global Negotiation stages. The outcome of a Side Deal—the Agreed Maximum Demand—applies only to Countries who took part in the Side Deal, though it is visible to all subjects in a group prior to the following Global Negotiation stages (see Figure 2 for details on the stages and Figure S.5-S.8 in SI for details on how the Agreed Maximum Demand is determined).

We employ an experimental design that allows for between-subject analysis. Each subject participated in a negotiation of up to eight rounds. Once all groups finished the negotiation, subjects were prompted to complete a brief questionnaire to assess motivation, strategic decision-making, and demographic heterogeneity (see SI; entire questionnaire available from the corresponding author upon request). Additionally, each subject answered a risk-preference elicitation question equivalent in structure to the standard question used in Eckel & Grossman (22; EG, hereafter), with payoffs scaled down to 10% of those used in EG. The question asked subjects to select one of five incentive-compatible 50-50 gamble options, where payoffs increase linearly in expected payout and "riskiness", as measured by the standard deviation of the two possible payouts (see also 23 and SI). The outcome of the gamble was determined individually by a coin toss upon payment for the study.

At the beginning of the experiment, subjects received both written and oral instructions (see SI). At the end of the experiment, subjects privately received their experimental earnings in cash, in addition to a £5 show-up fee, totaling £16.80 on average. All experimental decisions were made on a computer screen using the experimental software Z-Tree (24).

Subjects. A total of 336 student (undergraduate and postgraduate) and non-student subjects

volunteered to participate in 20 experimental sessions, most comprising three groups of six subjects (four sessions contained only two groups). The experiment took place at the London School of Economics (LSE), though experimental participation is not restricted to LSE students. In our sample, 50.9% of subjects were female, 42.3% are from developed countries (i.e. Annex 1; 5.2% USA, 36.3% EU, and 11.1% China), 47.6% are undergraduate students, and 33.6% are graduate students. The average age of our subjects is 23.5 years (SD=5.99). Participants come from various study disciplines (10.4% Business; 14.9% (International) Policy, Law, or Government; 8.0% Geography & Environment, 13.1% Economics).

RESULTS

Group-Level Success

i. Treatment Comparisons

Table 1 provides a descriptive overview of group performance across treatments. In terms of agreement velocity, the most successful treatment group is the one allowing for Side Deals among the Poor (PSD), where on average the groups coordinated on the threshold shortly after the second round. By contrast—RSD, where only the Rich engaged in preliminary side deals—was the treatment where agreement was most delayed (3.5 rounds on average). While ASYM and ASD are comparable in terms of the average agreement round, we note that there are two groups in the ASYM treatment who failed to reach agreement altogether, as indicated by the higher variance in agreement round for ASYM than for ASD. Similarly, while SYM and PSD are comparable along the former dimension, one group in the PSD treatment was not successful in coordinating on the threshold, consistent with the higher variance in outcomes for PSD than for SYM.

This variance is showcased in Table 2, where we see that *all* symmetric groups had reached agreement by the end of the fourth round of negotiations. That is, all symmetrically endowed groups maintained at least 50% of the initial pie, which is remarkably efficient given that the maximum attainable proportion of global wealth is 60%. When comparing success rates in Rounds 1-4, the SYM group outperforms ASYM (proportion test, $p=0.101$, $z=1.64$), RSD (proportion test, $p=0.049$, $z=1.96$), and ASD (proportion test, $p=0.062$, $z=1.86$) at conventional significance levels.

However, it is clear that the PSD treatment is the most successful in securing agreement under maximally efficient conditions, though we do not have the power to detect a statistically significant difference between groups' success rates within the first two periods. Regardless, if we are concerned with maximizing the global pie, endogenous demand restrictions on a larger number of "poor" players appear to be more successful in inducing coordination than similar restrictions on a smaller number of "rich" players. This outcome reinforces the simulation observation of Smead et al. (7). Importantly, we do not find conclusive evidence that treatments containing Side Deals improve upon global negotiations that occur among asymmetric actors in the absence of Side Deals, in terms of both agreement velocity and demands (at both the individual and the group level).

Figure 3 provides a visual representation of the above statistics in addition to the demand dynamics across treatments. The early disparity in agreement rate is clear, as is the tendency of average group demands to respond to the declining values of the Global Target T (from 60% to 30%) by clustering, although with some variance, around these values.

ii. Additional Determinants of Group Success

To provide additional insight into the determinants of group-level success, we look at differences between groups who were "successful"—which we define to be those groups who succeeded in reaching agreement without any efficiency losses (i.e. in Rounds 1 and 2)—and those who were not. Considering groups who participated in either the PSD or ASD treatments, the Agreed Maximum Demands of the Poor is not significantly different across successful and unsuccessful groups. However, if we look at groups in either the RSD or ASD treatments, the Agreed Maximum Demand of the Rich significantly differs across the successful and unsuccessful groups (WMW, 62.3 vs. 72.6, p=0.028, z=2.193). In fact, these differences hold if we compare these groups within RSD (WMW 58.4 vs. 66.6, p=0.106, z=1.616) and ASD (WMW, 65.5 vs. 78.6, p=.067, z=1.830) individually. This result indicates that the extent to which Rich Countries tie their hands is important for group success, though the same does not hold for Poor Countries.

We can also examine the effect of group composition in terms of proportion of individuals inclined to cooperate unconditionally (as opposed to those inclined to free ride), where unconditional cooperators are those who demand at most a percentage equivalent to the Global Target (T=60%) in Round 1. Pooling all treatments together, we find that there is almost exactly one additional unconditional cooperator on average in successful groups than in unsuccessful groups (WMW, 3.89 vs. 2.86, p=0.003, z=-2.945). This result remains when we exclude SYM from the comparison (WMW, 3.821 vs. 2.647, p=0.007, z=-2.703). We further investigate the importance of Rich vs. Poor cooperation and find that successful groups have almost double the number of Rich unconditional cooperators than unsuccessful groups, on average (WMW, 1.679 vs. 0.882, p=0.001, z=-3.426), while successful groups and unsuccessful groups are not significantly different in terms of the number of Poor unconditional cooperators (WMW, 2.14 vs. 1.76, p=0.400, z=-0.842). Taken together, these results indicate that strong commitment and unconditional cooperation by Rich Countries hold paramount influence in determining the success of multilateral negotiations.

Individual Demands

i. Poor vs. Rich

An interesting question pertains to the behavior of the two different player types in the asymmetric treatments: is there evidence of redistribution from the Rich to the Poor, in the sense of lower demands by the wealthy? If there are differences, do they persist over time? We tackle these questions in Figure 4. Interestingly, there are differences in initial demands, where the Poor Countries demand 66.7% of initial wealth while the Rich Countries demand 60.2% in the first round, on average (Wilcoxon-Mann-Whitney test, p=0.000, z=3.381). The 60% target appears to be salient for both groups across treatments—especially for the Rich in RSD and ASD—with the Rich not deviating far from it and the Poor demanding below ten percent above the threshold (see Figure 4 and Figure S9 in SI).

Even more interesting is the apparent dependence of this Poor-Rich disparity on whether Side Deals take place prior to the first global negotiation stage. Comparing the average initial demands of Poor and Rich Countries within treatment groups, we see substantial differences in the PSD (WMW, 67.3 vs. 57.8, p=0.071, z=1.805), RSD (WMW, 66.4 vs. 58.3, p=0.031, z=2.154), and ASD (WMW, 66.4 vs. 60.8, p=0.092, z=1.686) treatments, though this difference is attenuated in

the ASYM treatment (WMW, 66.7 vs. 62.9, p=0.240, 1.186). It thus appears that Side Deals increase the salience of the inequality, inciting self-serving fairness motivations that are manifested through a polarization of demands (15 and Figure S10 in SI).

This increased salience is especially apparent when the Side Deals pertain to only one subgroup (i.e. *either* the Poor *or* the Rich), as evidenced by the Maximum Demand inputs by Rich and Poor Countries in the various treatments containing side deals (see Figure S11 in SI). For instance, in PSD, the modal Maximum Demand input in the Side Deal pertaining to the first two rounds of Global Negotiation is 100%, and a vast majority of Poor Countries chooses values at or above the Global Target of 60%. On the contrary, in RSD, not a single player chooses 100% as her Maximum Demand, and a majority of Rich Countries selects a value between 46-70%. However, when both Poor and Rich Countries engage in Side Deals, the distribution of Maximum Demands between the two player types is strikingly similar.

Finally, the evolution of demands for Poor and Rich Countries differs. While the trend is negative for both types, the Poor display more variance across treatments, especially in the final rounds. The treatments where Poor subjects are most willing to reduce demands are RSD and ASD. Conversely, the opposite occurs for the Rich: PSD, followed by ASYM, is the treatment in which the Rich demand the least in the second half of the game, averaging only about 20% demand in the final round.[3] Note also that for the Poor, PSD is the treatment where demands are highest toward the end (close to 50% in Round 8), while for the Rich it is RSD that averages the highest demands in the final round (above 30%).

ii. Conditional Demands

The above observation raises further questions. Why it is that negotiators belonging to a given wealth group respond more cooperatively when the Side Deals take place in the other group? Is it a conditional cooperation argument, where the Poor (Rich) feel reassured about selflessly demanding less when the Rich (Poor) signal willingness to tie their hands in the upcoming negotiations?

Quite the opposite: we find a significant positive effect of past cooperation by the Rich on Poor Countries' demands, and similarly of past cooperation by the Poor on Rich Countries' demands (see Table 3). The Poor increase their demand in the present round by almost four percentage points on average for every additional Rich Country that cooperated (by demanding less than or equal to the target) in the previous round. Similarly, Rich increase their demands by almost three percentage points for each additional Poor Country that cooperated in the prior round. We do not find evidence that Countries free ride off of the cooperation of like Countries. In other words, Poor (Rich) Countries only free ride off of Rich (Poor) Countries' cooperation.

iii. Risk Preferences

To further understand the dynamics underpinning group coordination, we investigate the role of individual risk preferences in predicting behavior in the negotiation. As expected, we find that risk aversion reduces demand, on average. In Table 4, we display the effects of risk preferences on

[3] It is important to note that the sample size declines as negotiations progress, since groups who successfully reach agreement cease negotiating once agreement is reached.

individual demands. *Risk* is measured on a scale from 1 to 5, where 1 represents the most risk-averse gamble option—a gamble with payout certainty—and 5 represents the most risk-seeking option.

The table demonstrates that the effect of the risk parameter on demand is robust with respect to inclusion of various controls. The initial inclusion of controls—including demographics, stated motivation, Global Target, and treatment group assignment—reduces the magnitude of the effect from 1.68 to 1.24 percentage points per one-point increase on the risk scale. However, the magnitude of the effect is restored when we additionally account for the role of threshold (even) rounds—or rounds in which a failure to reach agreement results in negative consequences for the group— which have a large negative effect on demand, as expected.

We further investigate the role of threshold effects through the interaction term between threshold rounds and risk preferences. Since threshold rounds threaten to diminish global welfare, we expect risk-averse individuals to err on the side of caution by demanding less than risk-prone individuals in these rounds. In regression four, we see that the state of being in a threshold round reduces individual demand by almost six percentage points on average. However, the positive coefficient for the interaction term—which becomes significant when using self-reported risk preferences as the independent variable ($p=0.036$; see Table S3 in SI)—indicates that this threshold effect is less strongly negative the more risk seeking is the individual, as anticipated.

CONCLUSION

We explore the impact of wealth heterogeneity and minilateral agreements on climate bargaining processes in a laboratory setting. We do not find that "tying your hands" ahead of the inclusive negotiations promotes cooperation per se, although Side Deals among various subsets of players affect the cooperation dynamics differently. Specifically, Poor Side Deals on their own provide maximal agreement velocity, but treatments that include prior agreements among Rich players (i.e. Rich Side Deals and All Side Deals) experience success in all groups (contrary to the other treatments). Furthermore, the disparity between the average demand of Rich and Poor players widens in the presence of Side Deals, suggesting a larger role for Rich Countries. Finally, we find evidence that Poor and Rich Countries tend to free ride off of the other type's cooperation, and that risk-averse individuals demand less than risk-prone individuals, especially in rounds preceding a tightening of the target.

The above findings suggest that the infrastructure around which climate change negotiations revolve and the wealth and preferences of the players at the table may be crucial determinants of the dynamics behind climate negotiations. It appears that the dependence of the negotiation dynamics on the features of the process is largely driven by fairness concerns, as side deals among one negotiator type (e.g., among the Rich only) may increase the salience of inequitable wealth distributions. Future research should attempt to identify and explain the relevant motivations underlying contextual features of the bargaining process in order to understand the impacts of complex negotiation contexts and processes on bargaining outcomes.

What are the implications for the climate negotiations taking place in Paris in late 2015? Initial evidence from the Intended Nationally Determined Contributions pledged so far, mostly by high income countries, suggests that current ambitions are incompatible with the levels needed to limit

the change in mean global temperature to 2°C (6, 27). Furthermore, INDCs submitted so far by high income countries are highly heterogeneous, with Switzerland and the EU at one end of the spectrum (respectively -50% and -40% greenhouse gas emissions in 2030, relative to 1990), the US and Russia in the middle (approximately -25%), and Singapore and Republic of Korea at the other extreme (+112% and +86%, respectively). Our experiment sheds new light on the bargaining processes behind climate negotiations, revealing the importance of signaling for success. It shows that: (i) voluntary restrictions on future pathways, similar to INDCs, do not ensure greater success at the negotiation table; (ii) success crucially depends on the composition of the group, and particularly on the decisions of the Rich; (iii) a timely agreement thus hinges on strong commitment and unconditional cooperation by the wealthy.

Taken together, the above conclusions cast a shadow on the prospects for a sufficiently ambitious outcome in Paris. It appears that pledges are not "game changers", at least unless (iii) holds, which thus far has not been the case. To make matters worse, while the game analyzed here brings potentially disruptive wealth heterogeneities to center stage, other obstacles further hinder climate change cooperation. For instance, the game equates current emissions with responsibilities, neglecting historical responsibilities. Moreover, only six negotiators must strike an agreement, which simplifies the coordination problem. On the other hand, climate negotiations can rely on more instruments than those available to our subjects. Here there are no direct transfer mechanisms, such as the Adaptation Fund and climate finance. In addition, climate co-benefits may lure countries to join small-n clubs early on, providing much needed leadership (1-3). Our game focuses on short-run costs of mitigation, neglecting such opportunities. Yet, policy tends to be defined by short-term incentives and high discounting, as confirmed by the insufficient ambition of current INDCs (6, 27). Hence, we may be due for another inconclusive agreement in Paris, and we urge policymakers to consider additional complementary or stand-alone mechanisms to increase the chances of avoiding dangerous climate change.

ACKNOWLEDGEMENTS

The authors contributed equally to this work. Financial support by Enel Foundation (Grant 1-RGI-U885) is kindly acknowledged. This work is part of the program of the Centre for Climate Change Economics and Policy, which is funded by the UK Economic and Social Research Council. Special thanks to Jeroen Nieboer, Jamie Moss, and the Behavioural Research Lab for assisting with the software and the laboratory facilities.

REFERENCES

1. Keohane R, Victor D (2011) The regime complex for climate change. *Persp on Pol* 9(1): 7-23.
2. Ostrom E (2010) Nested externalities and polycentric institutions: must we wait for global solutions to climate change before taking actions at other scales? *Econ Theory* 49(2): 353-369.
3. Victor D (2006) Toward effective international cooperation on climate change: numbers, interests and institutions. *Global Environ Polit* 6(3): 90-103.

4. Barrett S (2010) A portfolio system of climate treaties. *Post-Kyoto International Climate Policy: Implementing Architectures for Agreement*, eds Aldy J and Stavins R (Cambridge U P, Cambridge), pp. 240-270.
5. Falkner R, Stephan H, Vogler J (2010) International climate policy after Copenhagen: towards a 'building blocks' approach. *Global Pol* 1(3): 252-262.
6. "INDCs as communicated by parties." <http://www4.unfccc.int/submissions/indc/Submission%20Pages/submissions.aspx>.
7. Smead R, Sandler R, Forber P, Basl J (2014) A bargaining game analysis of international climate negotiations. *Nature Climate Change* 4(6): 442-445.
8. Nash J (1950) The bargaining problem. *Econometrica* 18(2): 155-162.
9. Muthoo A (1999) *Bargaining theory with applications*. (Cambridge U P, New York).
10. Barrett S and Dannenberg A (2012) Climate negotiations under scientific uncertainty. *Proc Natl Acad Sci USA* 109(43): 17372-17376.
11. Barrett S, Dannenberg A (2013) Sensitivity of collective action to uncertainty about climate tipping points. *Nature Climate Change* 4(1): 36-39.
12. Dannenberg A, Löschel A, Paolacci G, Reif C, Tavoni A (2015) On the provision of public goods with probabilistic and ambiguous thresholds. *Environ Resource Econ* 61(3): 365-383.
13. Hasson R, Löfgren Å, Visser M (2010) Climate change in a public goods game: investment decision in mitigation versus adaptation. *Ecol Econ* 70(2): 331-338.
14. Milinski M, Sommerfeld R, Krambeck H, Reed F, Marotzke J (2008) The collective-risk social dilemma and the prevention of simulated dangerous climate change. *Proc Natl Acad Sci USA*, 105(7): 2291-2294.
15. Tavoni A, Dannenberg A, Kallis G, Loschel A (2011) Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc Natl Acad Sci USA* 108(29): 11825-11829.
16. Lenton T *et al.* (2008) Tipping elements in the Earth's climate system. *Proc Natl Acad Sci USA* 105(6): 1786-1793.
17. Pacheco J, Vasconcelos V, Santos F (2014) Climate change governance, cooperation and self-organization. *Phys Life Rev* 11(4): 573-586.
18. Santos F, Pacheco J (2011) Risk of collective failure provides an escape from the tragedy of the commons. *Proc Natl Acad Sci USA* 108(26): 10421-10425.
19. Steffen W *et al.* (2015) Planetary boundaries: guiding human development on a changing planet. *Science* 347(6223): 1259855-1259855.
20. Tavoni A (2013) Game theory: building up cooperation. *Nature Climate Change* 3(9): 782-783.
21. Vasconcelos V, Santos F, Pacheco J (2013) A bottom-up institutional approach to cooperative governance of risky commons. *Nature Climate Change* 3(9): 797-801.
22. Eckel C, Grossman P (2008) Forecasting risk attitudes: an experimental study using actual and forecast gamble choices. *J Econ Behav Organ* 68(1): 1-17.
23. Binswanger H (1980) Attitudes toward risk: experimental measurement in rural India. *Am J Agr Econ* 62(3): 395-407.
24. Fischbacher U (2007) z-Tree: Zurich toolbox for ready-made economic experiments. *Exp Econ* 10(2): 171-178.
25. Friedlingstein P *et al.* (2014) Persistent growth of CO_2 emissions and implications for reaching climate targets. *Nature Geoscience* 7: 709-715.
26. Libecap G (2014) Addressing global environmental externalities: transaction costs considerations. *J Econ Lit* 52(2): 424-479.

27. International Energy Agency (2015) World energy outlook special report: Executive summary, Paris.

FIGURES

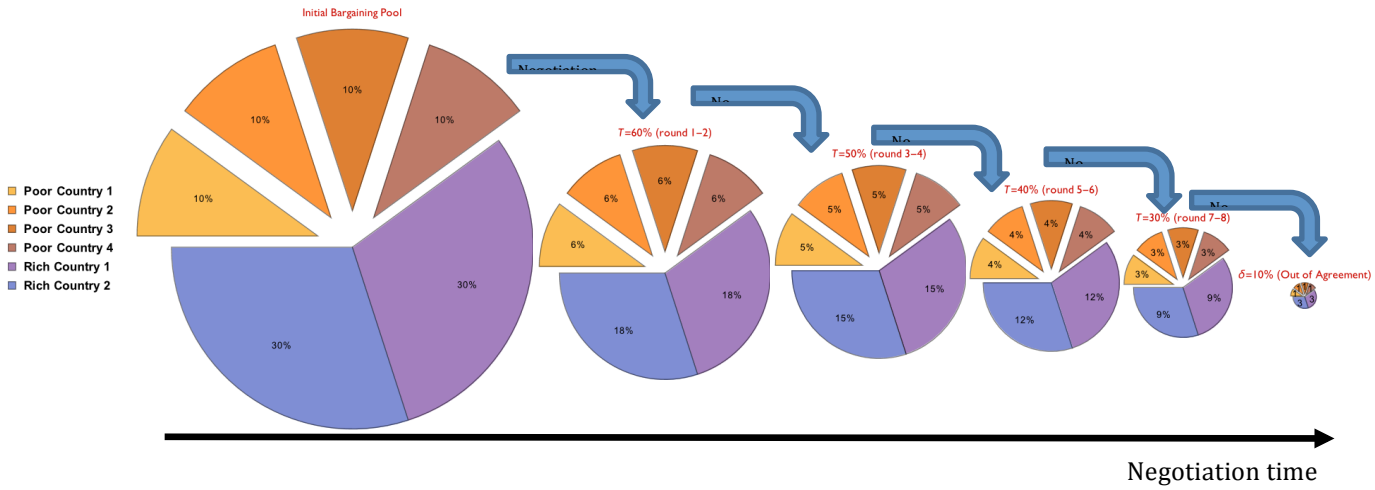


Figure 1. Timing and dynamics of the game. The six-player bargaining game begins with a collective "pie" of £100, which is split between two Rich Countries (each endowed with 30% of the pie, i.e. £30), and four Poor Countries (each endowed with 10% of the pie, i.e. £10). In the game, emissions map one-to-one onto wealth. That is, wealthier countries emit proportionally more CO_2 than poorer countries. Starting from this initial allocation of wealth/emissions, the group faces sequential rounds of bargaining on progressively tighter targets, from $T = 60\%$ in the first two rounds (i.e. a 40% reduction from BAU) to $T = 30\%$ in the final two rounds (i.e. a 70% reduction from BAU). Failure to agree on sufficient global reductions thus becomes more costly over time. The figure depicts the wealth/emissions distribution ensuing from each target if Countries were to reduce symmetrically; for instance, to keep with $T = 50\%$, Rich Countries would each have to halve their wealth to 15% of the pie, and Poor Countries would each have to halve it to 5% of the pie.

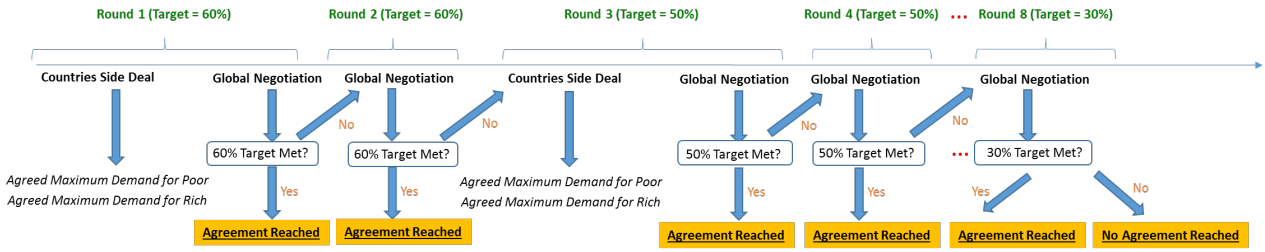


Figure 2. A schematic representation of the stages in treatment ASD. In even-numbered rounds there is only one stage (Global Negotiation), while in odd-numbered rounds that stage follows a Side Deal stage. The same applies to PSD and RSD, except that the Side Deal in those treatments are determined by (and pertain only to) Poor and Rich Countries, respectively.

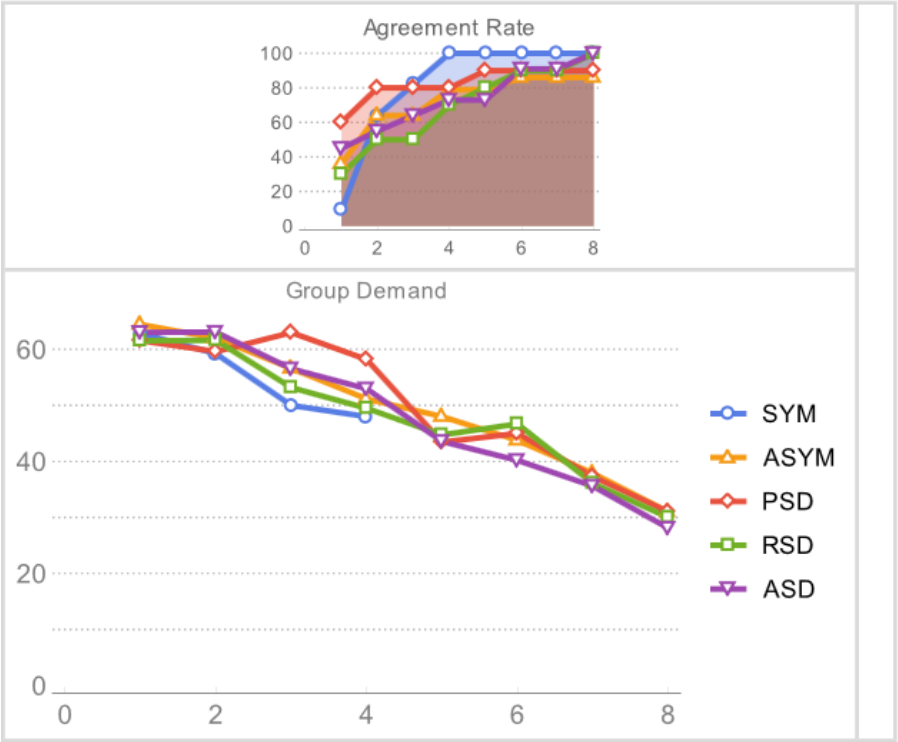


Figure 3. Group demand over time (and rate of agreement in the inset). The figure illustrates average group demands and agreement rates over time. The inset shows the percent of groups that reached agreement, by round and treatment. The main chart shows the average group demand, by round and treatment.

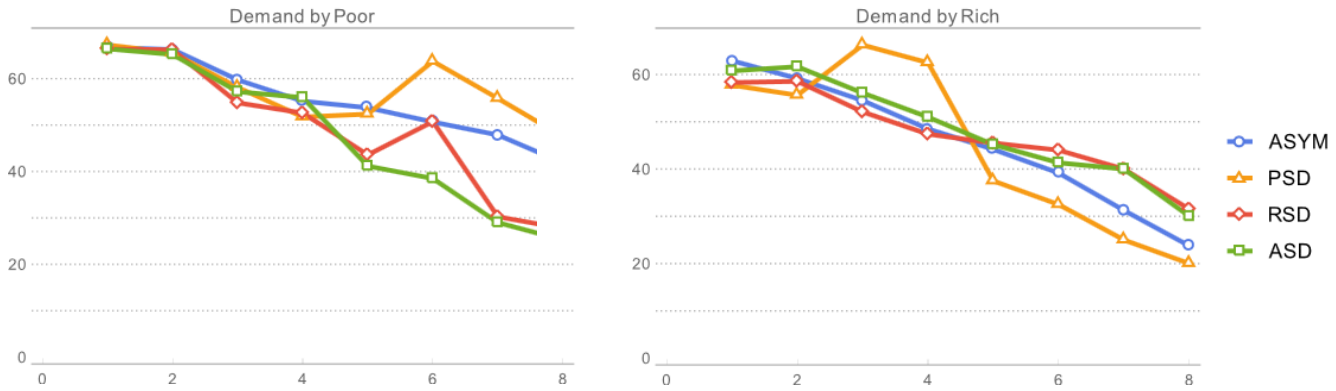


Figure 4. Demands over time by treatment, for the Poor (left panel) and for the Rich (right panel). The figure illustrates individual demands over time for both player types. The left panel plots the average demand of Poor Countries by round and treatment, and the right panel plots the average demand of Rich Countries by round and treatment.

TABLES

Table 1. Agreement velocity (average round in which negotiations terminated) and failures (number of groups that failed to reach an agreement), by treatment

	SYM	ASYM	PSD	RSD	ASD
Velocity	2.455	3.071	2.300	3.400	3.091
	(0.934)	(2.556)	(2.359)	(2.413)	(2.548)
Failures	0	2	1	0	0
Groups	11	14	10	10	11

Table 2. Success rate by target level

	Rounds 1-2	Rounds 1-4	Rounds 1-6	Rounds 1-8
SYM	63.64%	100.00%	100.00%	100.00%
ASYM	64.29%	78.57%	85.71%	85.71%
PSD	80.00%	80.00%	90.00%	90.00%
RSD	50.00%	70.00%	90.00%	100.00%
ASD	54.54%	72.73%	90.91%	100.00%

Table 3. Evidence of free riding by Poor and Rich

	Poor Demand	Rich Demand
Rich Cooperated	3.865**	1.694
	(1.768)	(2.540)
Poor Cooperated	-0.020	2.685***
	(1.047)	(0.813)
Constant	59.401***	53.175***
	(6.194)	(3.578)
Groups	26	26
Subjects	104	52
Obs	356	178
Controls	Yes	Yes

Notes: The dependent variable in this regression indicates the individual demands over the course of negotiation. The independent variables represent the number of Rich and Poor Country representatives (respectively) who cooperated in the prior round by demanding less than or equal to the Global Target. Controls include gender, Annex 1 nationality, stated primary motivation, Global Target, and the difference between the group demand and the target in the prior round of negotiations. There are 26 groups in heterogeneous treatments that negotiated past the first period, and these are the groups considered here. Robust errors are clustered at the group level. Standard errors are reported below estimates in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

Table 4. Risk preferences and individual demands

	(1) No Controls	(2) With Controls	(3) With Threshold Control	(4) With Threshold Interaction
Risk	1.680***	1.241**	1.769***	1.541**
	(0.535)	(0.562)	(0.666)	(0.678)
Threshold Round			-3.989***	-5.737***
			(0.550)	(1.015)
Threshold Round * Risk				0.476*
				(0.276)
Constant	52.198***	58.097***	61.459***	62.294***
	(2.270)	(2.631)	(3.105)	(3.069)
Groups	54	54	34	34
Subjects	324	324	204	204
Obs	930	930	810	810

Notes: The table displays the results of a panel OLS regression with errors clustered at the group level, where the dependent variable is individual demand. The risk question posed to subjects is based on the incentive-compatible risk preference elicitation gambles in (22, 23). Robust standard errors are reported in parentheses. *Threshold Round* is a dummy equal to 1 if the present round is the threshold round before a decline in the Global Target (i.e. an even round). The number of observations reduces with the threshold control since 20 groups who reach agreement in Round 1 will not experience variation in the *Threshold Round* control and are thus excluded from the regression. Controls include gender, Annex 1 nationality, stated primary motivation, Global Target, and treatment group assignment. *significant at 10%; **significant at 5%; ***significant at 1%.

Supporting Information

SI Materials and Methodology

Methodological Details. While the experimental literature on international climate change negotiations tends to center upon (framed) public goods games, we depart from this mechanism in several ways for two primary purposes: 1) to enhance the relevance of the context, and 2) to provide an empirical test of the agent-based model proposed in (1). Rather than employ a voluntary contribution mechanism devoid of context, we narrow our interest to pertain solely to international climate change negotiations, where the instructions provide clear background information on the economic complexities associated with this pervasive externality.

For instance, the dynamic nature of the Global Target captures the cost of delaying legislation to curb greenhouse gas emissions, a stock pollutant with long-term atmospheric effects. Furthermore, the composition of the groups—where a third of the countries represented are responsible for 60% of global greenhouse gas emissions—is reflective of the 54% for which the top three global players (United States, European Union, and China)—who have been engaging in pre-Paris minilateral discussions—are responsible (2).

Finally, the target persists over two rounds to allow for learning. While any time lapse theoretically increases the necessity of increased future abatement commitments to reach a given target (say, 2ºC), the slow and lagged process of climate change and the relative frequency of negotiations allows for fairly stable global goals in the short term, so that learning from one negotiation to the next is possible. Additional features of the game—e.g., the termination of negotiation if the target is reached, the detailed description of global climate change, and the correlation between emissions and wealth—were designed to mimic the climate context as closely as possible.

Our experimental design encapsulates features from similar experiments, namely (3). In their experiment, groups composed of six (asymmetrically endowed) players aim to avoid the losses associated with catastrophic climate change in a dynamic framed experiment. However, unlike (3), we do not impose a set number of rounds, and we do not vary the probability of climate catastrophe if the target is met. In our game, meeting the target guarantees payout, but payout is already associated with losses compared to the status quo (i.e. the initial endowment). Additionally, players in our asymmetric treatments received heterogeneous endowments, whereas those in (3) received symmetric endowments and a subset of players were 'forced' to contribute to a climate fund in the first three rounds to create asymmetry for the following seven rounds.

Experimental Instructions for participants of the ASD treatment. Welcome to the experiment! In this experiment, you can earn money. In addition to your earnings from the experiment, you will receive a £5 show-up fee. During the course of the experiment, please do not talk to other participants. We will now read the experimental instructions aloud. Once we have finished reading, raise your hand if you have questions and we will be with you shortly to answer them. At the end of Part A of the instructions you will find some questions that are meant to ensure that you understand the rules of the experiment. Please answer all questions and signal us by raising your hand when you have finished, so that we may check your answers.

Background: Climate change. Climate change is viewed as a serious global environmental problem. The vast majority of climate scientists expects the global average temperature to rise by 1.1-6.4°C before 2100, where a rise of 2°C is generally considered to be dangerous climate change. There is hardly any disagreement that mankind largely

contributes to climate change by emitting greenhouse gases, especially carbon dioxide (CO_2). CO_2 originates from the burning of fossil fuels such as coal, oil, or natural gas in industrial processes and energy production, as well as from combustion engines of cars and lorries. CO_2 is a *global* pollutant—that is, each unit of CO_2 emitted has the same effect on the climate regardless of the location where the emissions occur. Dangerous climate change will result in significant global costs, which get worse over time if agreement is not reached. International climate change negotiations involve yearly meetings where delegations representing different *countries try to strike a global agreement on emissions reductions* that are consistent with the goal of avoiding dangerous climate change. Here you will be asked to negotiate such costly emissions reductions on behalf of the Country to which you will be assigned. Your choices, together with those of the other 'Countries', will determine your payout from the experiment.

Rules of play. Now we will introduce you to a game simulating international climate change negotiations. In total, six Countries are involved in the global negotiation. That is, in addition to you, there are five other negotiators in your negotiation group, and each of you represents one Country. The six Countries account for *all* global wealth and CO_2 emissions (for simplicity, we disregard other greenhouse gases in the experiment). While excessive emissions impose global costs, individual Countries rely on productive processes which create emissions in order to generate wealth: for every 1 billion tons of CO_2 'emitted' in the game, you receive £1. Hence, reducing emissions is costly. Your decisions in the experiment are anonymous. To guarantee anonymity, you will be randomly assigned to one type of Country (Rich or Poor), and you will be identified by one of the following names: Rich Country 1, Rich Country 2, Poor Country 1, Poor Country 2, Poor Country 3, Poor Country 4. Your name will appear on the lower left side of your screen once the experiment begins. At the beginning of the experiment, you will receive a sum of money that represents your Country's wealth. This wealth mirrors your Country's CO_2 emissions. Therefore, throughout the instructions and the experiment, we will refer to wealth and emissions interchangeably. The current situation in your negotiation group can be summarised as follows:

- **Two Rich Countries** each emit 30 billion tons of CO_2 and earn **£30** in doing so;
- **Four Poor Countries** each emit 10 billion tons of CO_2 and earn **£10** in doing so;
- The resulting Global Emissions amount to **100 billion tons of CO_2** (2×30 billion tons of CO_2 + 4×10 billion tons of CO_2)
- Hence, **Global Wealth** is equal to **£100** (2×£30 + 4×£10)

Due to the threat of dangerous climate change, the goal is to agree on an aggregate level of Global Emissions that does not exceed a given Global Target. In the following experiment, you will participate in up to 8 rounds of climate change negotiations, where the global costs from *not* reaching agreement increase every 2 rounds. **Accordingly, the Global Target decreases every two rounds, as follows:**

- Rounds 1-2: **60% of current emissions** (60 billion tons of CO_2)
- Rounds 3-4: **50% of current emissions** (50 billion tons of CO_2)
- Rounds 5-6: **40% of current emissions** (40 billion tons of CO_2)
- Rounds 7-8: **30% of current emissions** (30 billion tons of CO_2)

To be clear, since current global emissions are 100 billion tons of CO_2, an agreement is only reached if total negotiated emissions are at most 60 billion tons of CO_2 in the first two rounds. Equivalently, Global Wealth must be reduced from an initial level of £100 to a target level of £60 if the Global Target is to be met in

the first two rounds. This target becomes more difficult to meet as the negotiations move forward, as outlined above. Every Country faces a similar decision-making problem. In each round of the global negotiation, all six Countries will be asked simultaneously: *"What percent of YOUR COUNTRY's emissions/wealth do you demand to keep?"* If the required Global Target is met, then your group has reached an agreement; negotiations terminate and each Country receives its demand from that round. If agreement is not reached, the negotiation continues to the next round. If an agreement is not reached by the end of the 8th Round of negotiations, dangerous climate change becomes unavoidable and economic costs for all Countries ensue. Each Country will then receive **10% of its initial wealth** (£3 for Rich Countries, £1 for Poor Countries).

Example 1. Imagine that you are part of a negotiation group that makes decisions as follows. In **Round 1** (Global Target=60%), all Countries demand to keep 90% of their emissions/wealth. If the Global Target were to be met, Rich Countries would receive £27 in payout and Poor Countries would receive £9 in payout. See **Figure S1** below, for the screen that will be seen by Poor Country 1. However, the Global Target is NOT met and negotiations continue to Round 2. In **Round 2** (Global Target=60%), demands are as follows:

Global Target
Rounds 1-2: **60%**
Rounds 3-4: **50%**
Rounds 5-6: **40%**
Rounds 7-8: **30%**

- Rich Country 1 and Poor Country 1 each demand to keep 50%. If the Global Target were to be met, Rich Country 1 would receive 50% of its initial wealth (£15) and Poor Country 1 would receive 50% of its initial wealth (£5).
- Rich Country 2 and all remaining Poor Countries (2,3,4) each demand to keep 80%. If the Global Target were to be met, Rich Country 2 would receive 80% of its initial wealth (£24) and Poor Countries 2, 3, and 4 would receive 80% of their initial wealth (£8 each).

See **Figure S2** below. However, Global Demand=68% > Global Target = 60%, so the Global Target is not met and negotiations continue. *Now imagine that the negotiation group continues to demand to keep emissions/wealth above the target level until the 7th Round, when the relevant Global Target is 30% of emissions/wealth.* In **Round 7**, demands are as follows:

- Rich Country 1 and Poor Country 4 demand to keep 32% each.
- Rich Country 2 and Poor Countries 1, 2, and 3 demand to keep 20% each.

See **Figure S3**.

Hence, Global Demand = 25% ≤ Global Target = 30%. The Global Target is met. Rich Country 1 receives 32% of its initial wealth (£9.60), Rich Country 2 receives 20% of its initial wealth (£6), Poor Countries 1, 2, and 3 each receive 20% of their initial wealth (£2 each), and Poor Country 4 receives 32% of its initial wealth (£3.20). Please take a brief moment to review and understand the rules, then continue to the next page to test your understanding.

***Control questions.* Test your understanding:** *For the questions below, please check the box of the correct answer or fill in your answer on the line provided. For convenience, we summarised the main rules below:*

Country Initial Wealth
Rich Country 1, Rich Country 2: **£30**
Poor Country 1, Poor Country 2, Poor Country 3, Poor Country 4: **£10**

1. In Round 4's global negotiation, all members of your negotiation group demand to keep 60% of their initial emissions/wealth. What happens next?

- ☐ *We've met our Global Target; each of us receives 60% of our initial wealth.*
- ☐ *Our Global Target has not been met; we continue to Round 5.*

2. In Round 3's global negotiation, all Rich Countries demand to keep 50% of their original emissions/wealth. If two Poor Countries demand to keep 40% and the other two Poor Countries demand to keep 60%, is agreement reached?

- ☐ Yes
- ☐ No

If yes, how much does each Country receive (without show-up fee)? If no, please leave blank.

Rich Countries: £_____each

Poor Countries that demanded 60%: £_____each

Poor Countries that demanded 40%: £_____each

3. In the final Round's global negotiation (i.e. Round 8), one Rich Country demands to keep 20% of its initial emissions/wealth and the other Rich Country demands to keep 30%. If two Poor Countries demand to keep 30% each and the other two Poor Countries demand to keep 75% each, is agreement reached?

- ☐ Yes
- ☐ No

How much does each Country receive as their final payout (without show-up fee)?

Rich Country that demanded 20%: £_____

Rich Country that demanded 30%: £_____

Poor Countries that demanded 30%: £_____each

Poor Countries that demanded 75%: £_____each

Please raise your hand when you have answered all questions, and we will come to check your answers.

Side Deals. Recall that the Global Target changes every two rounds. Before global negotiations on a new target begin, both groups of Countries (the 4 Poor and the 2 Rich) will simultaneously enter into separate side deals, as follows.

(i) Side Deal for Poor Countries:

Prior to the global negotiations in Rounds 1, 3, 5, and 7, each Poor Country will enter its preferred 'Maximum Demand', i.e. the desired maximum percentage of emissions/wealth that each *Poor Country* may demand to keep in the two upcoming global negotiations.

The average of these four Maximum Demands will determine the 'Agreed Maximum Demand for Poor', which cannot be exceeded by each Poor Country in the two upcoming global negotiations.

(ii) Side Deal for Rich Countries:

At the same time, and prior to the global negotiations in Rounds 1, 3, 5, and 7, each Rich Country will enter its preferred 'Maximum Demand', i.e. the desired maximum percentage of emissions/wealth that each *Rich Country* may demand to keep in the two upcoming global negotiations.

The average of these two Maximum Demands will determine the 'Agreed Maximum Demand for Rich', which cannot be exceeded by each Rich Country in the two upcoming global negotiations.

Should a global agreement *not* be reached within the first two rounds, a new target will apply to Round 3 (Global Target=50%) and a

new Agreed Maximum Demand will be determined by both Poor and Rich Countries for the two upcoming rounds (Rounds 3 and 4). This process will continue until Round 8 so long as a global agreement is not reached. Please refer to the timeline in **Figure S4** for a recap on the various stages of the game.

Example 2. Imagine that you are Poor Country 1 and that you have entered into a side deal with the other Poor Countries. In the experiment you will see the following screen (**Figure S5**).

The choices from the Side Deal for Poor Countries are shown at the top of **Figure S6**, which we have highlighted with a box:

- Poor Country 1 (you) chooses Maximum Demand = 100%
- Poor Country 2 chooses Maximum Demand = 66%
- Poor Country 3 chooses Maximum Demand = 33%
- Poor Country 4 chooses Maximum Demand = 0%

The resulting agreed side deal is that each Poor Country cannot exceed 50% demand in the two upcoming global negotiations, i.e. the Agreed Maximum Demand = 50%. (Note that the outcomes of the Side Deal for Rich Countries, which took place at the same time, are also shown in **Figure S6**. All Countries see these outcomes.)

Example 3. Imagine that you are Rich Country 1 and that you have entered into a side deal with Rich Country 2. In the experiment you will see the following screen (**Figure S7**).

The choices from the Side Deal for Rich Countries are shown at the bottom of **Figure S8**, which we have highlighted with a box:

- Rich Country 1 (you) chooses Maximum Demand = 75%
- Rich Country 2 chooses Maximum Demand = 25%

The resulting agreed side deal is that each Rich Country cannot exceed 50% demand in the two upcoming global negotiations, i.e. the Agreed Maximum Demand = 50%.

Questionnaire. Immediately following the experiment, subjects were asked a series of questions to gather demographic information, preferences (i.e. for fairness, risk, environment), and motivations in the experiment.

We look at players' primary decision-making motivations, acknowledging that the *ex post* nature of the questionnaire may create dependence of answers on dynamics and outcomes of the game played previously. When asked what is the most important motivation behind their decisions in the game, most claimed to have been primarily motivated by group efficiency (36.3% of subjects) or money (29.1% of subjects), with fairness (19.6% of subjects) following close behind. The rest were motivated by time minimization (7.5% of subjects), beliefs about actual climate negotiations (3.6% of subjects), and the past behavior of group members (3.6% of subjects). If money was a subjects' primary motivation, she initially demanded 6.9 percentage points more than if her primary motivation were not money. Additionally, we asked subjects what is the minimum demand they would accept if they were a pivotal player in the final round of negotiation, i.e. when the Global Target is 30%. The average minimum acceptable demand is 30.2 percent (SD=16.8) of one's endowment, and this is not largely dependent on whether one was a Poor (mean=31.9, SD=17.2) or Rich (mean=28.5, SD=17.1) player.

We also ask a series of questions to elicit our subjects' risk and social preferences. Using a variant of the incentive compatible risk preference elicitation question posed in (22, 23)—where 1 represents a certain outcome (50% chance of £1 vs. 50% chance of £1) and 5 represents the most risky outcome (50% chance of £4.20 and 50% chance of -£0.60)—subjects' average selection is 3.77 (SD=1.33). When asked to self-assess the extent to which

they are risk prone on a scale from 0 to 10 (where 10 is extremely risk prone), subjects' average selection is 5.38 (SD=2.11). To assess subjects' altruism, we ask them to state the frequency with which they donate to charity: 6.9% of subjects give to charity *very often*, 17.7% given to charity *often*, 45.8% give to charity *sometimes*, 28.1% give to charity *rarely*, and 1.6% *never* give. We also asked subjects the extent to which they think others can be trusted on a scale from 1 (low trust) to 10 (high trust), and the mean response is 4.9 (SD=2.1). To get a reading of subjects' preferences for the environment, we asked how often the subjects recycle. In our pool, 27.4% claim to recycle *very often*, 39.0% recycle *often*, 19.1% recycle *sometimes*, 4.8% recycle *rarely*, and 9.8% *never* recycle. Additionally, when asked their opinion on the severity of the problem of climate change, 35.0% of subjects responded that it is *extremely serious*, 36.0% that it is *very serious*, 19.6% that it is *serious*, 8.2% that it is *somewhat serious*, and 1.3% that it is *not at all serious*. Group-level heterogeneity in self-reported charitable spending and 'green' preferences do not play a significant role in determining subjects' decision-making nor the velocity of agreement in the game, contrary to the assertion that heterogeneity of preferences increases the transaction costs associated with (and therefore decreases the likelihood of) reaching agreement (26).

To gauge whether subjects understood the experiment, we ask the extent to which the instructions are comprehensible and ask for an evaluation of subjects' own ability to work with fractions. Subjects appear to have understood the experiment, with only six subjects (i.e. less than two percent) stating that the experiment was (very) difficult to understand. Similarly, only 2.3% of subjects claim they are *not at all good* with fractions, while 9.2% are *somewhat good* with fractions, 27.8% are *good* with fractions, 37.3% are *very good* with fractions, and 23.5% are *extremely good* with fractions.

Additional Empirical Analysis: Self-Serving Bias. Our data allows for empirical estimation of self-serving bias (4, 5). In the questionnaire described above, subjects were asked a series of survey questions, one of which pertained to their perspectives on equity in the context of climate change. To test for self-serving bias, we look at the average marginal effects of logit regressions where the dependent variables are dummies for whether the particular equity perspective in question has been selected, and the independent variables are indicators for subjects' nationality (United States, European Union, or China). Controlling for whether subjects played the role of a Rich or Poor Country in the experiment, we find that European subjects were less likely to identify with the perspective that "Countries with high emissions in the past should reduce more emissions" by 12.95% (p=0.038), although they were somewhat more likely than non-Europeans to identify with the perspective that "Countries with high economic performance should reduce more emissions" by 9.7% (p=.123). Additionally, we find that Chinese subjects were less likely to select "Countries with high economic performance should reduce more emissions" by 15.64% (p=0.055). We do not find definitive evidence of self-serving bias among Americans in our sample; however, American subjects were less likely to identify with the perspective that "Countries should reduce their emissions in such a way that emissions per capita are the same for all countries" than non-Americans by 13.8%, though the effect is not quite significant at conventional levels (p=0.140).

Supporting Analysis (Robustness). To account for the maximum demand imposed in the experimental design, we run panel Tobit regressions (see Tables S1 and S2) to complement the panel OLS regressions included in the manuscript (see Tables 3 and 4). This regression places an upper limit of 100 on individual demands. Since subjects may wish to demand more than 100 percent of

their endowed share of global emissions, the Tobit regressions censor the dependent variable from above at 100. Note that it is not necessary to censor the dependent variable from below since none of the experimental subjects demanded zero emissions in the game. The results of the Tobit regressions align closely with those of the OLS regressions in the manuscript, providing a simple robustness check of the free riding result and the influence of risk preferences on individual demands.

We run an additional panel OLS regression (see Table S3), replacing the incentive compatible risk preference with a stated preference for risk as our dependent variable. Again, the results are qualitatively similar to those in Table 4 of the manuscript. While the incentive compatible risk responses map preferences on a scale from 1 to 5, the stated risk responses map preferences on a scale from 0 to 10. Standard errors are slightly inflated relative to the OLS regression on the incentive compatible risk preference. However, the results for Regressions 1-3 are qualitatively similar. Interestingly, the results for Regression 4 indicate a positive though non-significant effect of risk preference on demand in the game, while the interaction between threshold round and risk becomes significant. That is, subjects who state a higher risk tolerance demand more in threshold rounds than do those who report less risk tolerance.

References

1. Smead R, Sandler R, Forber P, Basl J (2014) A bargaining game analysis of international climate negotiations. *Nature Climate Change* 4(6): 442-445.
2. Friedlingstein P *et al.* (2014) Persistent growth of CO_2 emissions and implications for reaching climate targets. *Nature Geoscience* 7: 709-715.
3. Tavoni A, Dannenberg A, Kallis G, Loschel A (2011) Inequality, communication, and the avoidance of disastrous climate change in a public goods game. *Proc Natl Acad Sci USA* 108(29): 11825-11829.
4. Lange A, Vogt C (2003) Cooperation in international environmental negotiations due to a preference for equity. *Journal of Public Economics* 87:2049-2067.
5. Lange A, Löschel A, Vogt C, Ziegler A (2010) On the self-interested use of equity in international climate negotiations. *European Economic Review* 54:359-375.Libecap G (2014) Addressing global environmental externalities: transaction costs considerations. *J Econ Lit* 52(2): 424-479.
6. Eckel C, Grossman P (2008) Forecasting risk attitudes: an experimental study using actual and forecast gamble choices. *J Econ Behav Organ* 68(1): 1-17.
7. Binswanger H (1980) Attitudes toward risk: experimental measurement in rural India. *Am J Agr Econ* 62(3): 395-407.
8. Dohmen T, Falk A (2011) Performance pay and multidimensional sorting: productivity, preferences, and gender. *Am Econ Rev* 101(2): 556-590.

FIGURES IN EXPERIMENTAL INSTRUCTIONS: Screenshots from Ztree

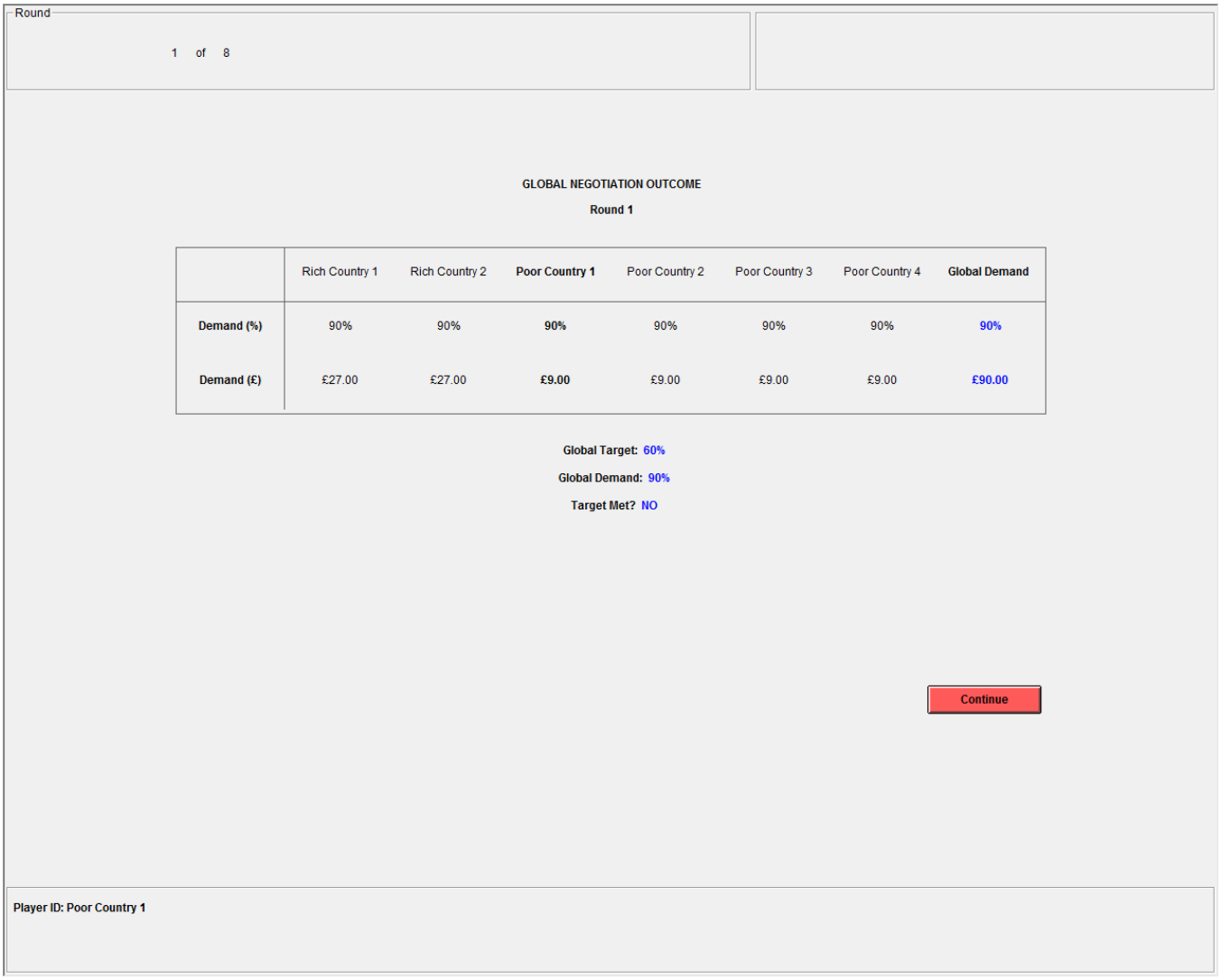


Figure S1. Outcome screen presented to Poor Country 1 in Round 1 of the Global Negotiation if all group members demand to keep 90% of their initial wealth/emissions in ASYM, PSD, RSD, and ASD. The Global Demand exceeds the Global Target of 60% in Round 1 and negotiations continue to Round 2.

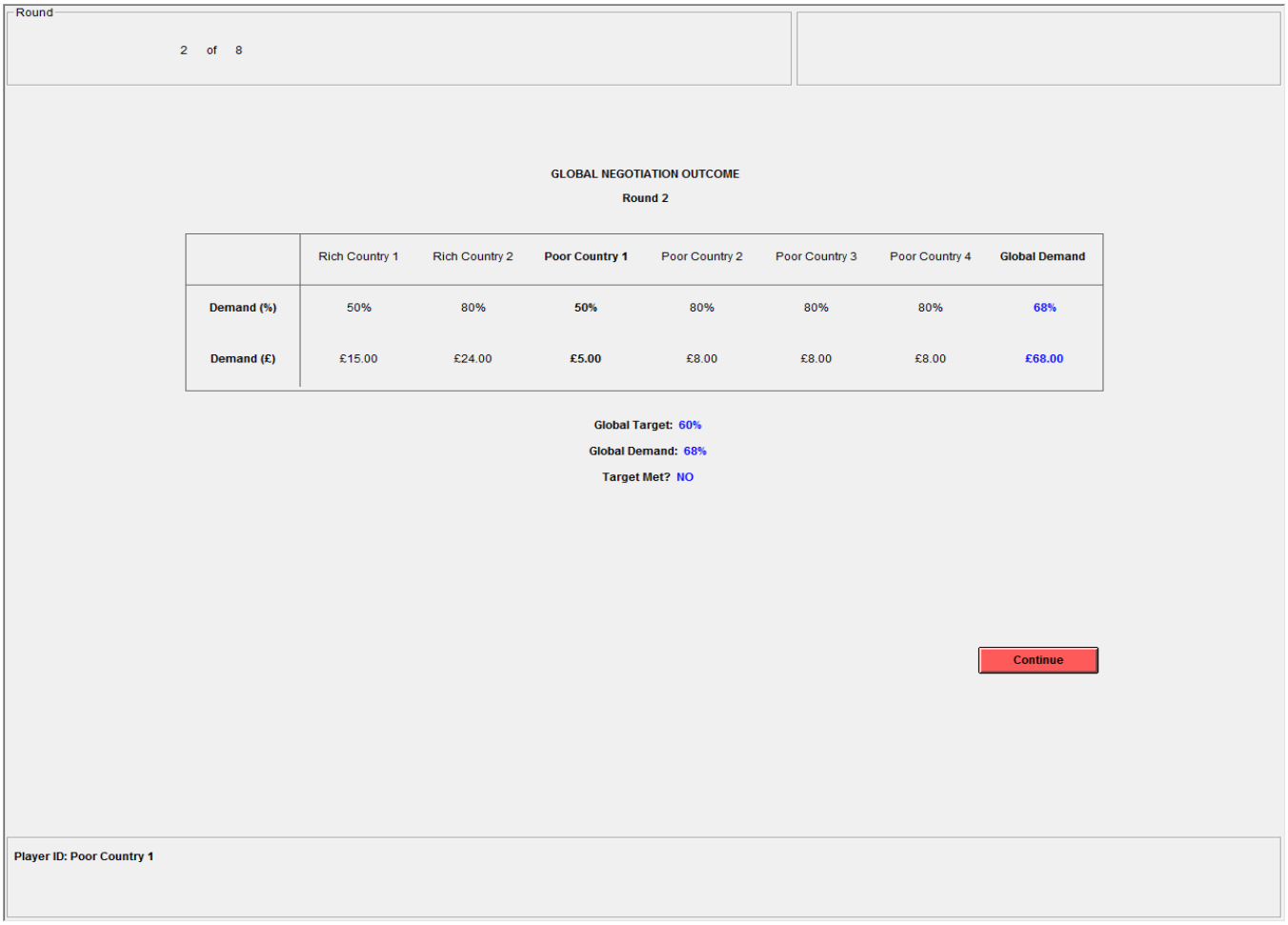


Figure S2. Outcome screen presented to Poor Country 1 in Round 2 of the Global Negotiation if Rich Country 1 and Poor Country 1 demand to keep 50% of their initial wealth/emissions and all other players demand to keep 80% of their initial wealth/emissions. The Global Demand exceeds the Global Target of 60% in Round 1 and negotiations continue to Round 3.

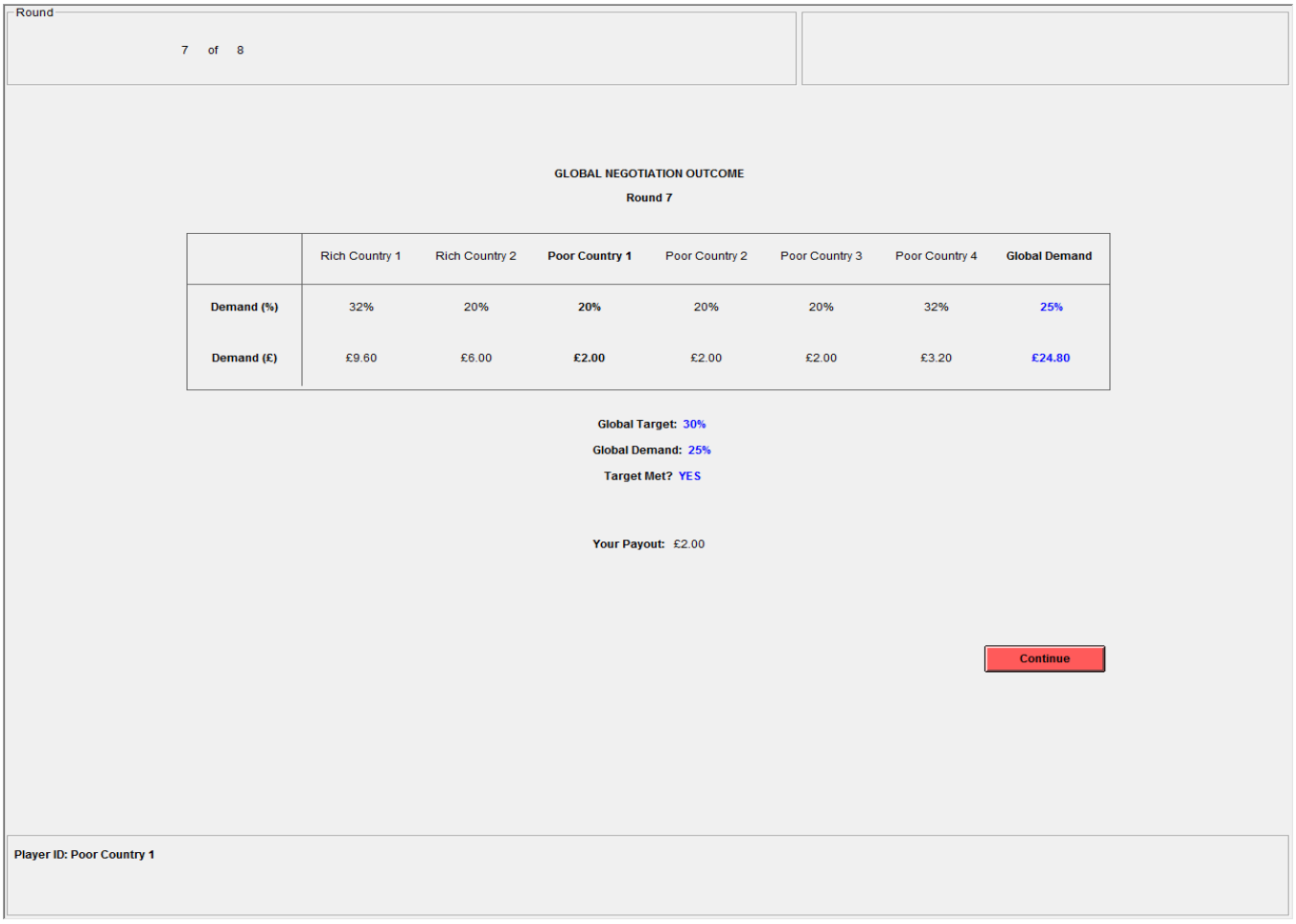


Figure S3. Outcome screen presented to Poor Country 1 in Round 7 of the Global Negotiation if Rich Country 1 and Poor Country 4 demand to keep 32% of their initial wealth/emissions and all other players demand to keep 20% of their initial wealth/emissions. The Global Demand is less than the Global Target of 30% in Round 7. Each player receives her demand and negotiations terminate.

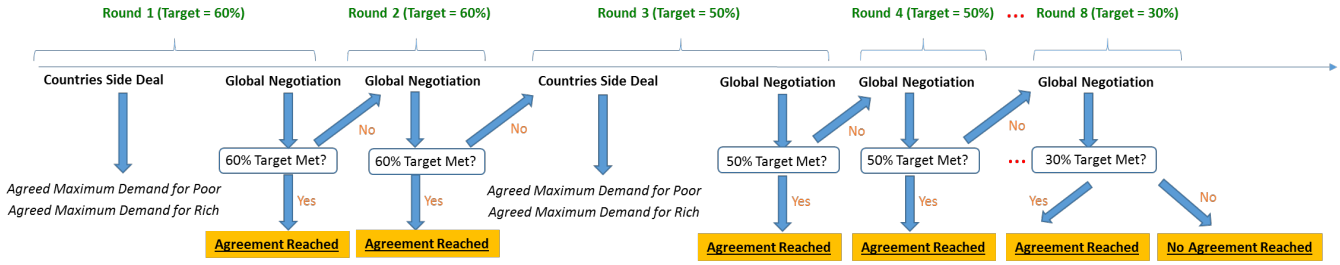


Figure S4. A timeline of each phase of the negotiation in the ASD treatment.

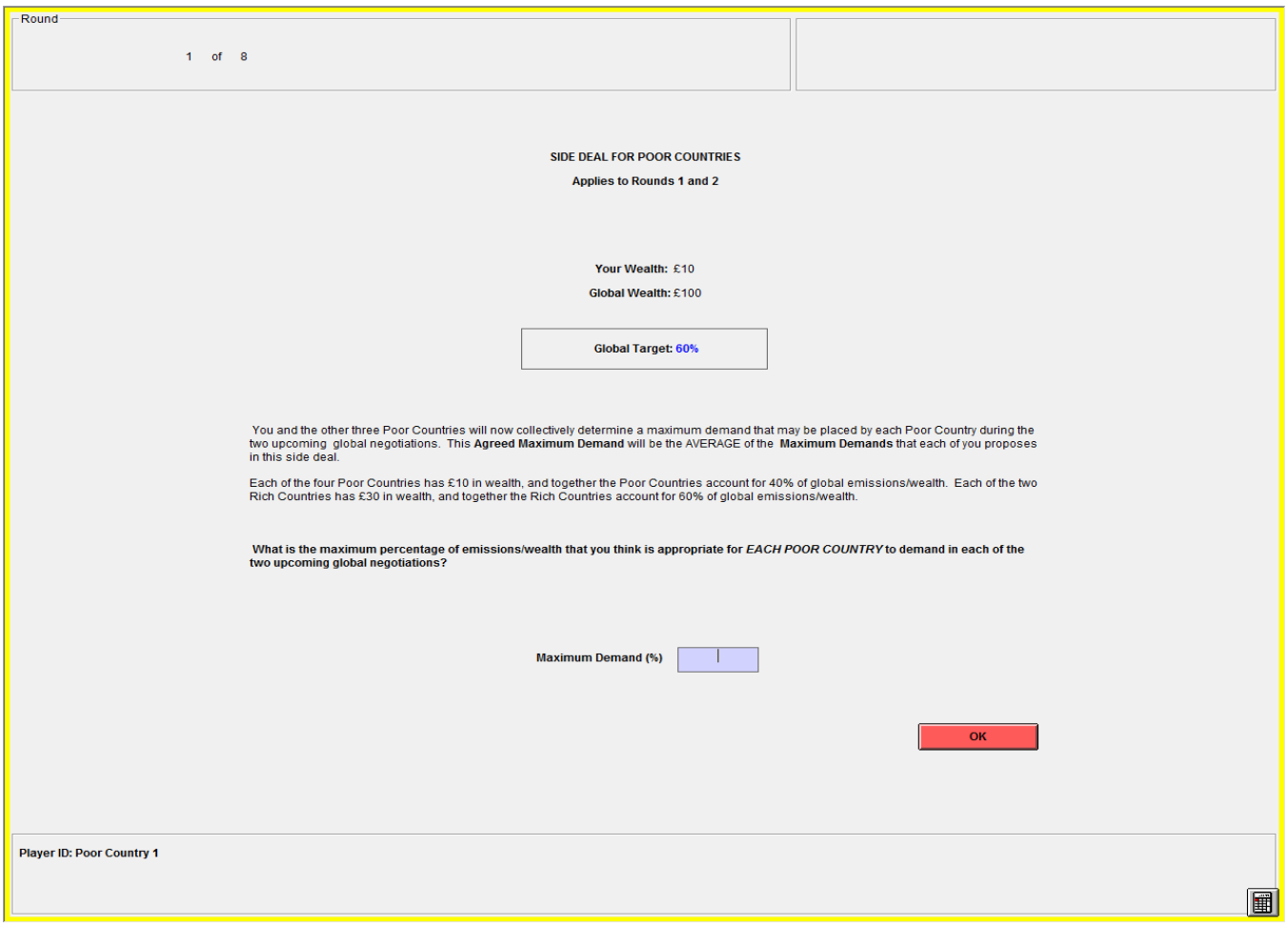
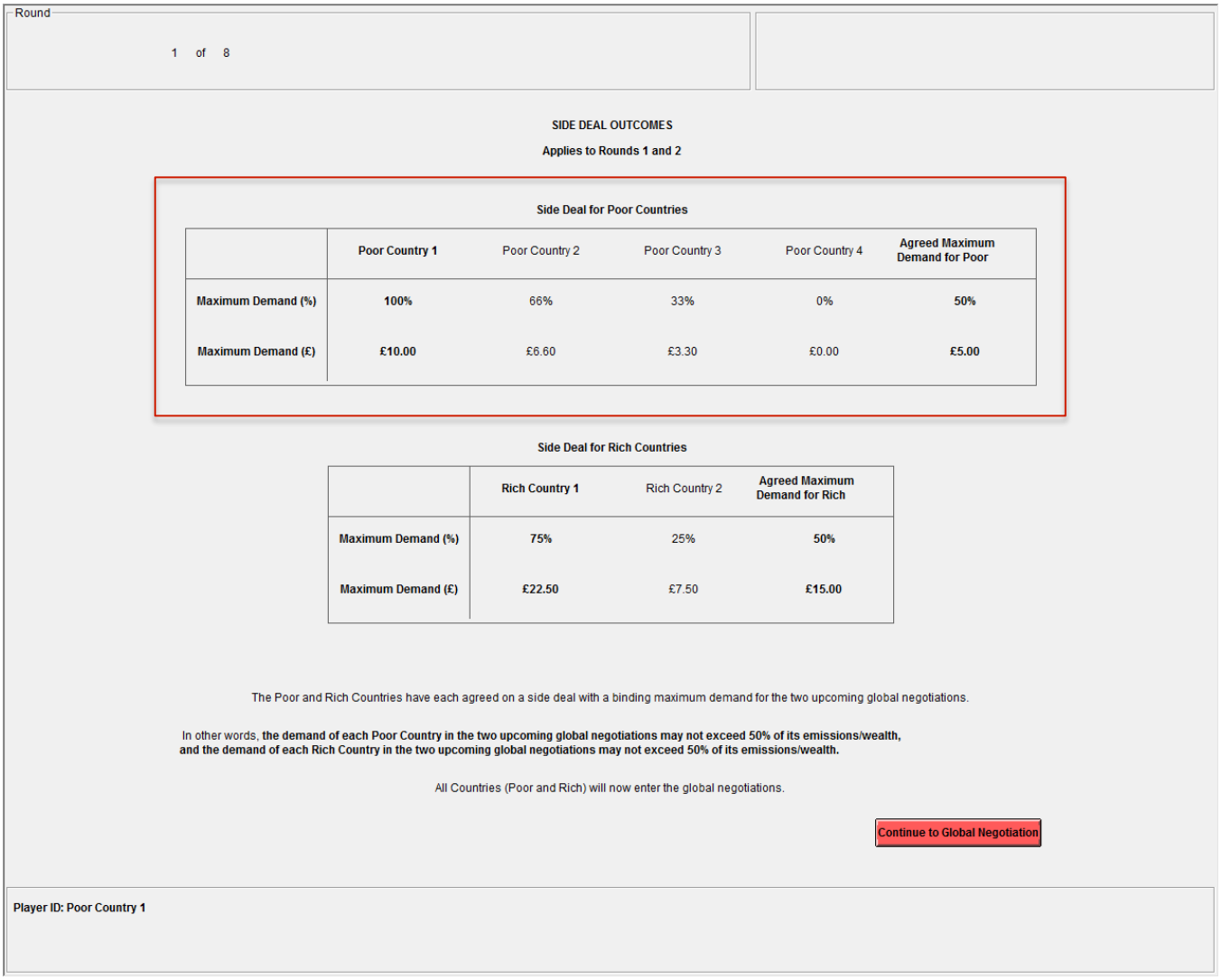


Figure S5. Input screen presented to Poor Country 1 to designate a preferred Maximum Demand in the Poor Countries' Side Deal prior to Rounds 1 and 2 of the Global Negotiation.



Round

1 of 8

SIDE DEAL OUTCOMES

Applies to Rounds 1 and 2

Side Deal for Poor Countries

	Poor Country 1	Poor Country 2	Poor Country 3	Poor Country 4	Agreed Maximum Demand for Poor
Maximum Demand (%)	100%	66%	33%	0%	50%
Maximum Demand (£)	£10.00	£6.60	£3.30	£0.00	£5.00

Side Deal for Rich Countries

	Rich Country 1	Rich Country 2	Agreed Maximum Demand for Rich
Maximum Demand (%)	75%	25%	50%
Maximum Demand (£)	£22.50	£7.50	£15.00

The Poor and Rich Countries have each agreed on a side deal with a binding maximum demand for the two upcoming global negotiations.

In other words, **the demand of each Poor Country in the two upcoming global negotiations may not exceed 50% of its emissions/wealth, and the demand of each Rich Country in the two upcoming global negotiations may not exceed 50% of its emissions/wealth.**

All Countries (Poor and Rich) will now enter the global negotiations.

Continue to Global Negotiation

Player ID: Poor Country 1

Figure S6. Outcome screen presented to Poor Country 1 displaying the selected Maximum Demands of all other players in her group. The red box is included in the Experimental Instructions to highlight the relevant Agreed Maximum Demand from the perspective of Rich Country 1, though it does not appear on screen during the experiment. Agreed Maximum Demands for both Rich Countries and Poor Countries are revealed to all group members prior to the subsequent Global Negotiation stages.

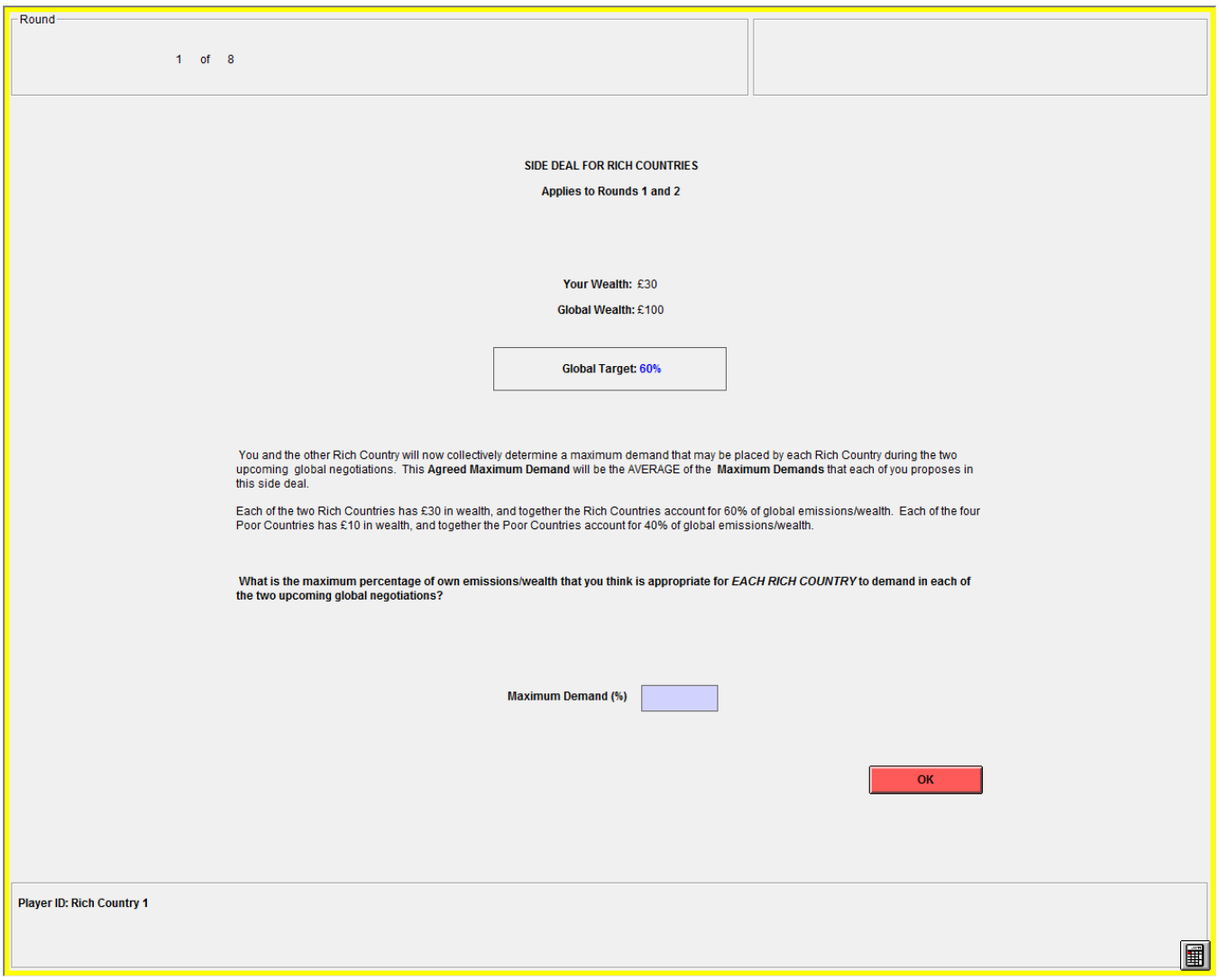
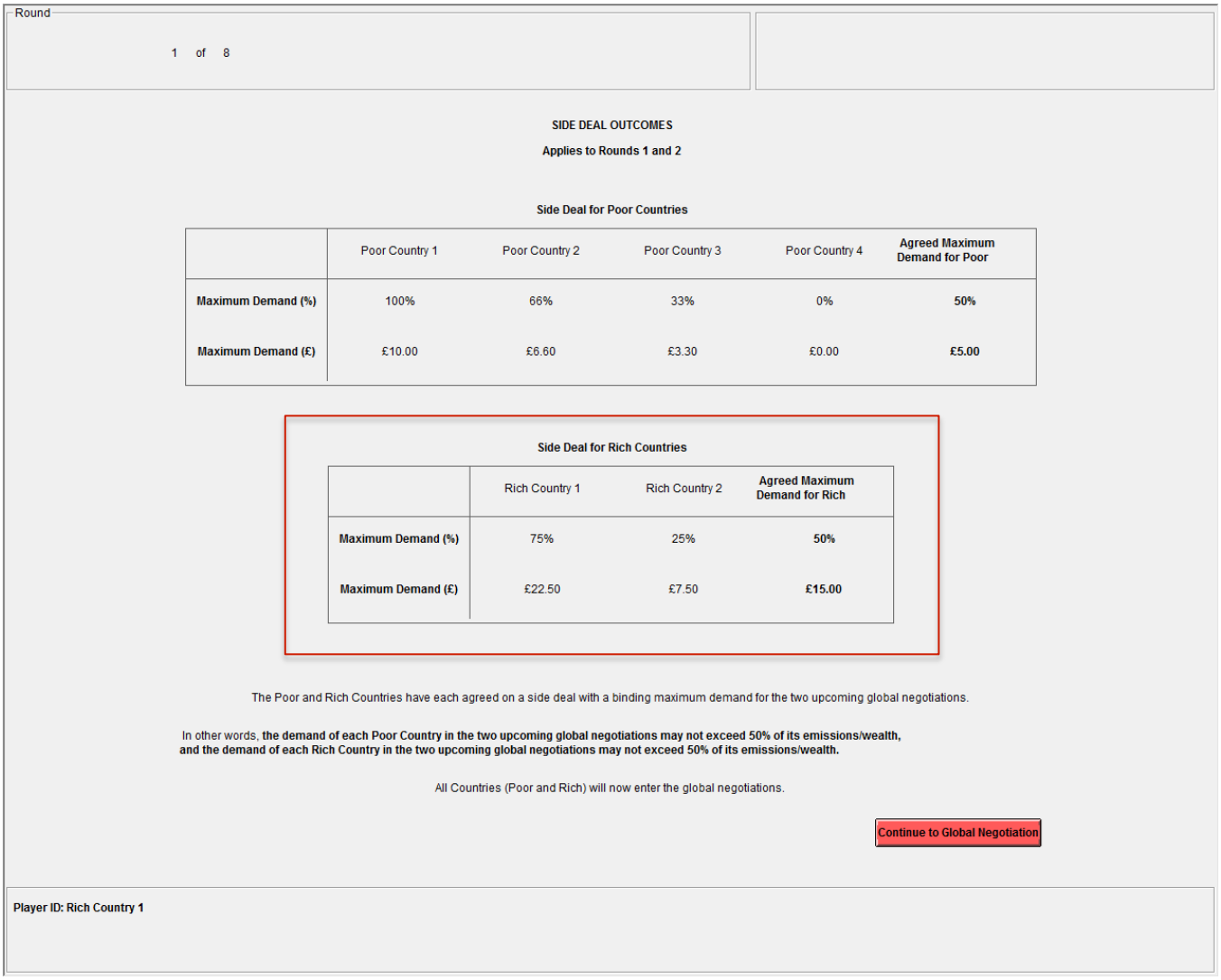


Figure S7. Input screen presented to Rich Country 1 to designate a preferred Maximum Demand in the Rich Countries' Side Deal prior to Rounds 1 and 2 of the Global Negotiation.



Round

1 of 8

SIDE DEAL OUTCOMES

Applies to Rounds 1 and 2

Side Deal for Poor Countries

	Poor Country 1	Poor Country 2	Poor Country 3	Poor Country 4	Agreed Maximum Demand for Poor
Maximum Demand (%)	100%	66%	33%	0%	50%
Maximum Demand (£)	£10.00	£6.60	£3.30	£0.00	£5.00

Side Deal for Rich Countries

	Rich Country 1	Rich Country 2	Agreed Maximum Demand for Rich
Maximum Demand (%)	75%	25%	50%
Maximum Demand (£)	£22.50	£7.50	£15.00

The Poor and Rich Countries have each agreed on a side deal with a binding maximum demand for the two upcoming global negotiations.

In other words, **the demand of each Poor Country in the two upcoming global negotiations may not exceed 50% of its emissions/wealth, and the demand of each Rich Country in the two upcoming global negotiations may not exceed 50% of its emissions/wealth.**

All Countries (Poor and Rich) will now enter the global negotiations.

Continue to Global Negotiation

Player ID: Rich Country 1

Figure S8. Outcome screen presented to Rich Country 1 displaying the selected Maximum Demands of all other players in her group. The red box is included in the Experimental Instructions to highlight the relevant Agreed Maximum Demand from the perspective of Rich Country 1, though it does not appear on screen during the experiment. Agreed Maximum Demands for both Rich Countries and Poor Countries are revealed to all group members prior to the subsequent Global Negotiation stages.

ADDITIONAL FIGURES

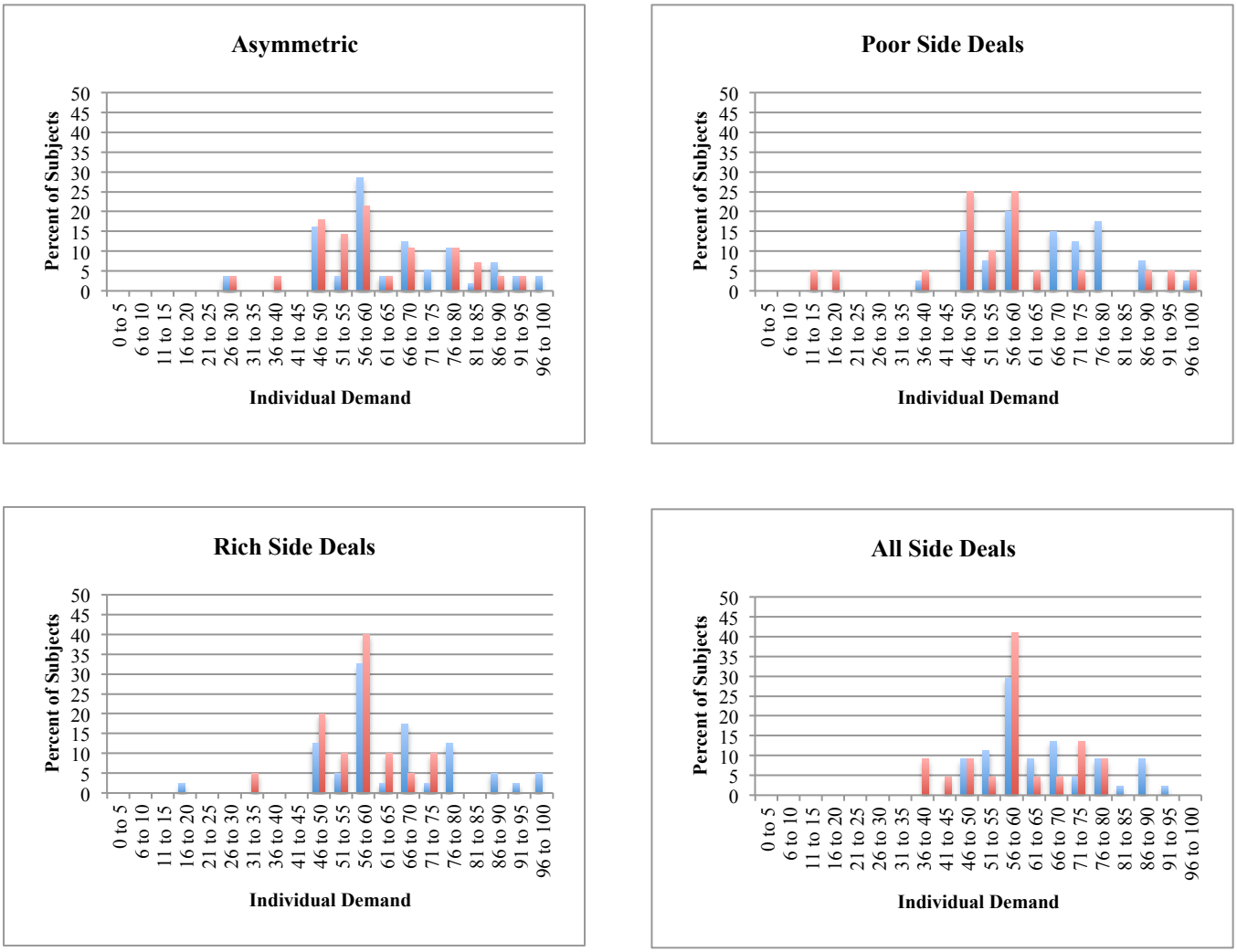


Figure S9. Distributions of individuals demands across Poor (blue) and Rich (red) negotiators in treatments with asymmetric endowments.

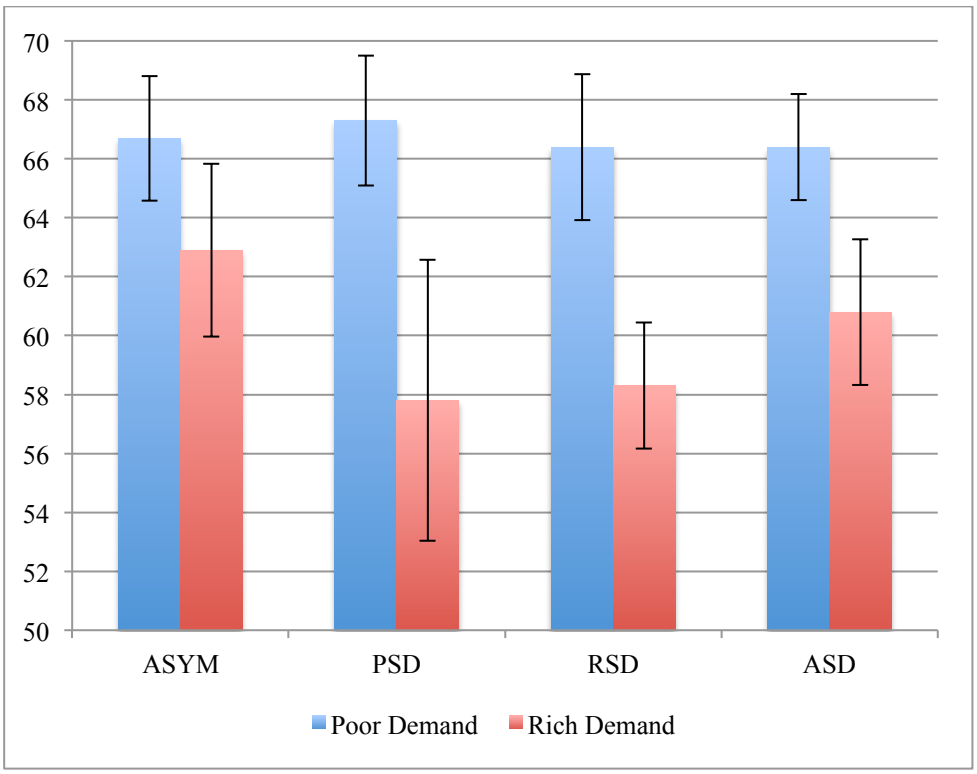


Figure S10. Average demands (and error bars) by Poor (blue) and Rich (red) negotiators in treatments with asymmetric endowments.

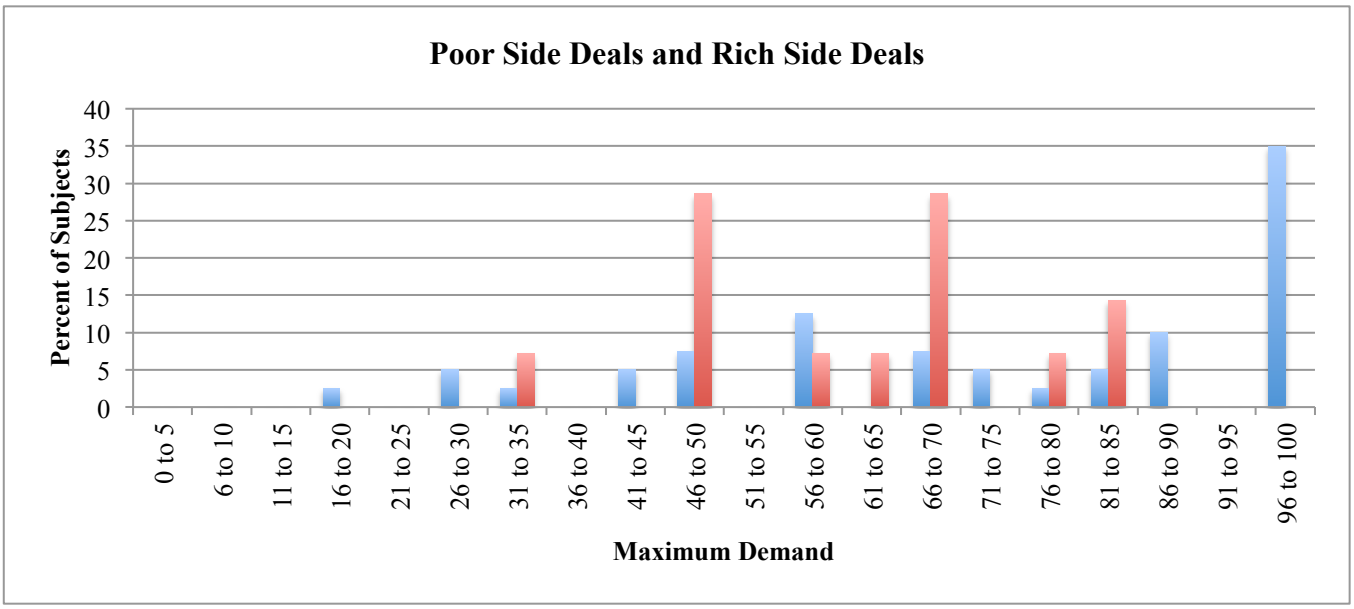


Figure S11. Distributions of Maximum Demands by Poor (blue) and Rich (red) players in treatments with Side Deals. Since only Poor (Rich) Countries input Maximum Demands in the Poor (Rich) Side Deals treatment, the top graph combines the data from these two treatments for ease of comparison.

TABLES

Table S1. Evidence of free riding by Poor and Rich (Tobit)

	Poor Demand	Rich Demand
Rich Cooperated	4.074***	0.766
	(1.301)	(2.024)
Poor Cooperated	-0.265	2.420***
	(0.740)	(0.805)
Constant	59.397***	55.995***
	(4.464)	(5.315)
Groups	26	26
Subjects	104	52
Obs	356	178
Controls	Yes	Yes

The table displays the results of a panel Tobit regression, where the dependent variable indicates the percentage demanded of one's initial endowment. The independent variables represent the number of Rich and Poor Country representatives (respectively) who cooperated in the prior round by demanding less than or equal to the Global Target. Controls include gender, Annex 1 nationality, stated primary motivation, Global Target, and the difference between the group demand and the target in the prior round of negotiations. There are 26 groups in heterogeneous treatments that negotiated past the first period, and these are the groups considered here. Robust errors are clustered at the group level. Standard errors are reported below estimates in parentheses. *significant at 10%; **significant at 5%; ***significant at 1%.

Table S2. Risk preferences and individual demands

	(1) Risk	(2) With Controls	(3) With Threshold Interaction	(4) With Threshold Interaction
Risk	1.659***	1.269**	1.805***	1.575**
	(0.536)	(0.541)	(0.604)	(0.635)
Threshold Round			-4.015***	-5.775***
			(0.563)	(1.613)
Threshold Round * Risk				0.480
				(0.412)
Constant	52.429***	58.142***	61.541***	62.382***
	(2.135)	(2.910)	(3.340)	(3.417)
Groups	54	54	34	34
Subjects	324	324	204	204
Observations	930	930	810	810

The table displays the results of a panel Tobit regression, where the dependent variable is individual demand. The risk question posed to subjects is based on the incentive-compatible risk preference elicitation gambles in 6, 7). Robust standard errors are reported in parentheses. *Threshold Round* is a dummy equal to 1 if the present round is the threshold round before a decline in the Global Target (i.e. an even round). The number of observations reduces with the threshold control since 18 groups who reach agreement in Round 1 will not experience variation in the *Threshold Round* control and are thus excluded from the regression. Controls include gender, Annex 1 nationality, stated primary motivation, Global Target level, and treatment group assignment. *significant at 10%; **significant at 5%; ***significant at 1%.

Table S3. Stated risk preferences and individual demands

	(1) No Controls	(2) With Controls	(3) With Threshold Control	(4) With Threshold Interaction
Stated Risk	0.697**	0.728**	0.673*	0.433
	(0.313)	(0.305)	(0.355)	(0.350)
Threshold Round			-3.987***	-6.703***
			(0.550)	(1.505)
TR * Stated Risk				0.496**
				(0.236)
Constant	54.702***	54.510***	64.652***	65.954***
	(1.811)	(2.195)	(2.746)	(2.667)
Groups	54	54	34	34
Subjects	324	324	204	204
Obs	930	930	810	810

The dependent variable in this regression is individual demand. Stated Risk is measured on a scale from 0 to 10 and comes from the general risk question asked in the German Socioeconomic Panel (SOEP; see 8). Robust standard errors are reported in parentheses. Controls include gender, Annex 1 nationality, stated primary motivation, Global Target level, and treatment group assignment. *significant at 10%; **significant at 5%; ***significant at 1%.